Machine Learning-Based Classification of Mental Health State Using the DASS-21 Profile

Conference Paper · November 2024

DOI: 10.1109/BMEiCON64021.2024.10896316

3 authors:

Khanita Duangchaemkarn
University of Phayao

Patiphan Khammarew
University of Phayao

Supavadee Aramvith
Chulalongkorn University

Machine Learning-Based Classification of Mental Health State Using the DASS-21 Profile

Khanita Duangchaemkarn*
Department of Pharmacy Practice,
School of Pharmaceutical Sciences,
University of Phayao, Phayao, 56000,
Thailand (Corresponding email:
kkhanita.du@up.ac.th)

Patiphan Khammarew
Pharmacy research extern under the
Doctor of Pharmacy Program,
School of Pharmaceutical Sciences,
University of Phayao, Phayao, 56000,
Thailand

Supavadee Aramvith
Multimedia Data Analytic and
Processing Research Unit, Department
of Electrical Engineering, Faculty of
Engineering Chulalongkorn University,
Bangkok, 10330, Thailand

***Abstract*— This study aimed to determine the most effective machine learning model for classifying emotional states of depression, anxiety, and stress based on the DASS-21 profile. Six machine learning models were developed and tested using 2,602 records from the University of Phayao Counselling Center, collected between 23 July and 16 August 2020. The results demonstrate that the deep learning model provided the highest accuracy and precision, exceeding 90% for all three emotional categories, with an AUC greater than 99%. These findings introduce an innovative approach for classifying emotional states of depression, anxiety, and stress using the DASS-21 profile as a benchmark. Incorporating additional features into the model could further enhance its utility in clinical decision-making, improving the accuracy of mental health screening and diagnosis.**

***Index Terms*— Machine Learning, Mental health, Depression, Classification.**

I. Introduction

Mental health problem, especially depression, has become a major concern since the first year of the COVID-19 pandemic. The prevalence of anxiety and depression rapidly increased by 25% globally, according to World Health Organization report [1]. There are multiple stress factors related to the university students, such as the rapid change from onsite to the online classroom, an isolated situation due to pandemics, less face-to-face activity, and, most important, the family's economic which directly affect their mental health status and may lead to depression, anxiety, or stress [2]. Mental health among university students is a major concern in many countries because this group is entering its early professional careers, which is the most important population in developing nations.

Depression Anxiety Stress Scales (DASS) is a scoring-based self-assessment questionnaire widely used for pre-screening depression, anxiety, and stress. DASS-42, first developed by Lovibond et al. (1995), is the gold standard Likert-like scale questionnaire. DASS-21 and DASS-12 are the short versions of the gold standard, which has 21 or 12 questions, but the assessment result remains in high accuracy when compared to DASS-42 [3], [4]. Marta et al. [5] performed the confirmatory factor analysis for DASS-42, DASS-21, and DASS-12. The result showed that Cronbach's alpha coefficient for DASS-42, DASS-21, and DASS-12 are 0.96, 0.93, and 0.89, respectively. There were attempts to apply machine learning models to predict the severity of depression, anxiety, and stress using the DASS-42 profile. In Priya et al. [6], the prediction results revealed that Naïve Bayes gave the best accuracy for all depression, anxiety, and stress at 0.855, 0.733, and 0.742, respectively. Prince et al. [7] reported that using a multilayer perceptron has performed best with the accuracy of depression, anxiety, and stress at 0.931, 0.988, and 0.965, respectively.

Translating the machine learning techniques to replace the conventional DASS-21 calculation, will make DASS-21 more robust and intelligent. The ML-based DASS-21 can be implemented into other smart health systems or redefined with extra features of interest. However, there is limited research on developing the machine learning-based model from the DASS-21 profile. This study aims to identify the state-of-the-art machine learning-based model to classify the emotional states of depression, anxiety, and stress. The findings reported in this study will benefit healthcare professionals in translating the model into a clinical setting which will improve clinical decision-making in mental health state evaluation and diagnosis.

II. Methods

A. Data Collection and Profiling

To build the DASS-21 profile for machine learning-based model prediction, 2,624 records were retrieved from the University of Phayao Counselling Center (UPCC) database which completed the online DASS-21 questionnaire from 23 July to 16 August 2020 [8]. The dataset was de-identification from the UPCC to protect personal data privacy following the Personal Data Protection Act 2019 [9] before transferring the dataset for analysis. From 2,624 records, fourteen missing data and 9 invalid data were removed. A total of 2,602 records were used for analysis. The demographic characteristic of the study population is described in Table 1.

B. Prediction models

Six standard machine learning classification techniques were chosen to develop the prediction models using Python Scikit-learn 1.2.1 [10]. The techniques used in this study consist of deep learning, k-nearest neighbor (k-NN), support vector machine (SVM), random forest (RF), Naïve Bayes, and adaptive boosting (AdaBoost). For deep learning, we constructed the deep neural network (21,50,50,50,3) using an MLP-based algorithm with rectified linear unit function (ReLu) and stochastic gradient-based optimizer (Adam) as a solver for

979-8-3315-0543-1/24/$31.00 ©2024 IEEE

weight optimization (α=0.0001, maximum iteration is 200). For k-NN, the estimated number of neighbors is 20, and the measuring method is Euclidean distance. For SVM, the penalty cost was set to 1,00, and the regression loss epsilon (ε) was set to 0,10. The RBF kernel was selected with γ = 0.047. For random forests, the number of trees has been set to 10. For AdaBoost, the number of estimators is 50, with a learning rate of 1.00. The SAMME classification algorithm was used as a boosting method with a linear regression loss function.

TABLE I. DEMOGRAPHIC CHARACTERISTICS OF THE STUDY POPULATION

Characteristics	*N (%)*
Gender	
- Female	1682 (64.10)
- Male	874 (33.31)
- Not specified	68 (2.59)
Field of study	
- Liberal arts and Social sciences	587 (22.37)
- Life Science and Technology	802 (30.56) 1235
- Health Science and Medicine	(47.07)
Under student loan debt	1334 (50.84)
Year	
- Freshman	421 (16.04)
- Second year	683 (26.03)
- Third year	718 (27.36)
- Senior	712 (27.13)
- Extern	90 (3.43)

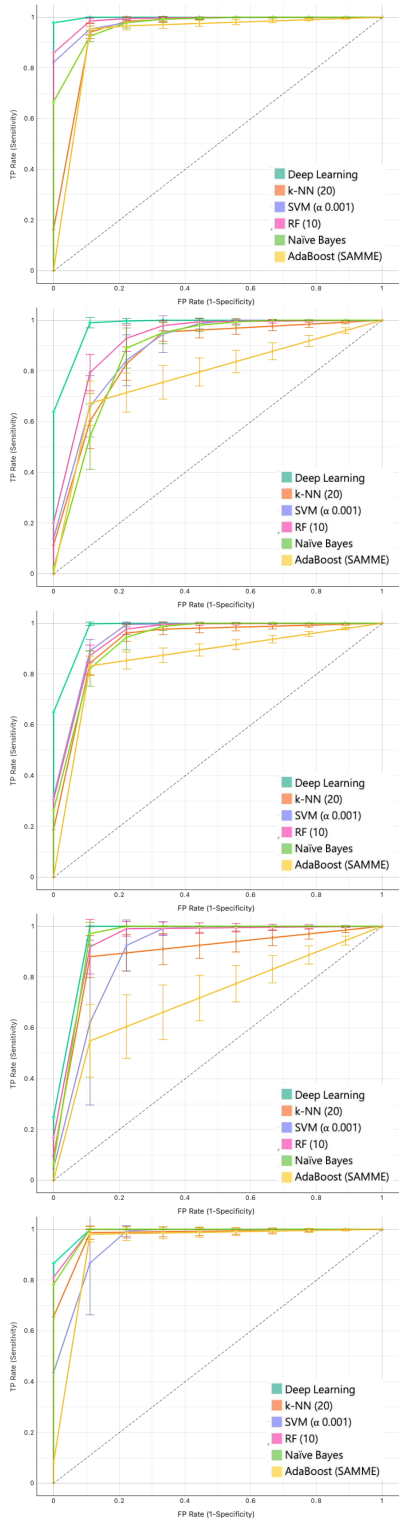
To build the model, each question's answers were used (21 attributes from 21 questions). The models were trained to predict the emotional state level of depression, anxiety, and stress. The predicted output is stratified into five severity levels: normal, mild, moderate, severe, and extremely severe—the results from the scoring-based technique, as shown in Table 2. The actual results will be compared with the predicted results performed by the developed machine learning models. The performance evaluation for each model is reported as the area under the curve (AUC), classification accuracy (CA), F1, precision, recall, and ROC analysis based on 10-fold cross-validation.

TABLE II. THE DEMOGRAPHIC OF DASS-21 ASSESSMENT RESULTS

	Depression	*Anxiety*	*Stress*
Normal	1652 (62.96)	1852 (70.58)	1988 (75.76)
Mild	309 (11.78)	275 (10.48)	261 (9.95)
Moderate	416 (15.85)	226 (8.61)	178 (6.78)
Severe	105 (4.00)	85 (3.24)	153 (5.83)
Extremely severe	142 (5.41)	186 (7.09)	44 (1.68)

III. RESULTS

In this study, six machine learning-based models were evaluated for classifying the emotional states of depression, anxiety, and stress using the DASS-21 profile. The model performance evaluation reported in ROC analysis, AUC, and F1 which are the measurement standards to compare the performance of the model with imbalance samples.



Category: Depression
Class: 'Normal'
N = 1652

Model	Evaluation	
	AUC	F1
DL	1.000	0.993
k-NN	0.974	0.919
SVM	0.985	0.961
RF	0.994	0.958
Bayes	0.976	0.912
AdaBoost	0.938	0.958

Category: Depression
Class: 'Mild'
N = 309

Model	Evaluation	
	AUC	F1
DL	0.995	0.910
k-NN	0.879	0.405
SVM	0.904	0.654
RF	0.944	0.620
Bayes	0.879	0.493
AdaBoost	0.801	0.651

Category: Depression
Class: 'Moderate'
N = 416

Model	Evaluation	
	AUC	F1
DL	0.995	0.914
k-NN	0.938	0.656
SVM	0.962	0.685
RF	0.967	0.765
Bayes	0.943	0.670
AdaBoost	0.889	0.759

Category: Depression
Class: 'Severe'
N = 105

Model	Evaluation	
	AUC	F1
DL	0.991	0.741
k-NN	0.912	0.143
SVM	0.894	0.961
RF	0.959	0.488
Bayes	0.970	0.563
AdaBoost	0.740	0.480

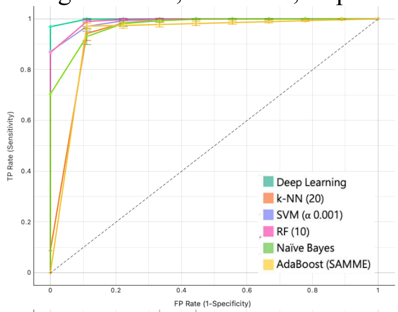
Category: Depression
Class: 'Extremely Severe'
N = 142

Model	Evaluation	
	AUC	F1
DL	0.999	0.931
k-NN	0.989	0.830
SVM	0.965	0.595
RF	0.998	0.883
Bayes	0.996	0.870
AdaBoost	0.982	0.856

Figure 1. The ROC analysis of depression's emotional state

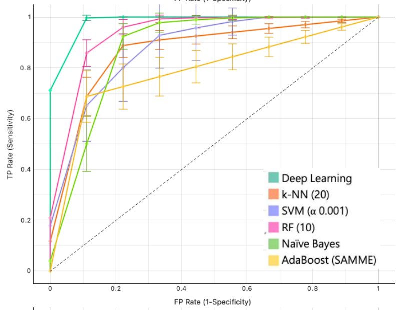
From Fig.1, the results show that deep learning has the highest AUC and F1 in every class. Especially with the 'Normal' class which has the highest number of samples. The average AUC and F1 from all classes are 0.996, and 0.951, respectively. The ROC results for the anxiety emotional state and stress emotional state also reflect the same trend.

From Fig.2, the ROC analysis in the ‘severe’ class is not unified because of the lowest sample size which reflects the F1 of deep learning to be as low as 0.595, even though the AUC is 0.982. However, the average AUC and F1 from all classes are still high at 0.996, and 0.956, respectively.



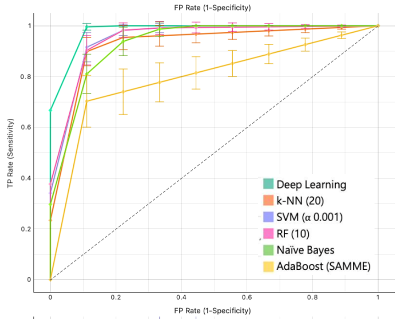
Category: Anxiety
Class: ‘Normal’
N = 1852

Model	Evaluation	
	AUC	F1
DL	0.999	0.989
k-NN	0.968	0.936
SVM	0.991	0.971
RF	0.993	0.961
Bayes	0.978	0.912
AdaBoost	0.949	0.971



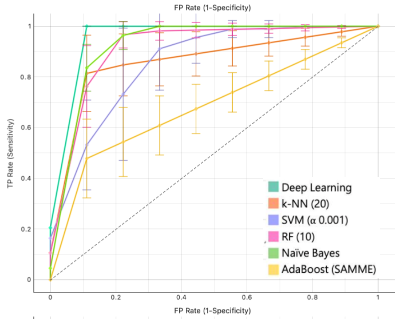
Category: Anxiety
Class: ‘Mild’
N = 275

Model	Evaluation	
	AUC	F1
DL	0.995	0.892
k-NN	0.878	0.311
SVM	0.890	0.638
RF	0.946	0.560
Bayes	0.880	0.452
AdaBoost	0.809	0.647



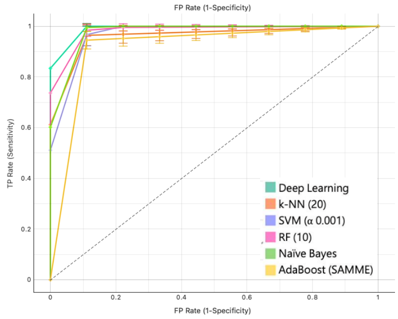
Category: Anxiety
Class: ‘Moderate’
N = 226

Model	Evaluation	
	AUC	F1
DL	0.994	0.858
k-NN	0.939	0.613
SVM	0.963	0.635
RF	0.957	0.626
Bayes	0.942	0.583
AdaBoost	0.821	0.643



Category: Anxiety
Class: ‘Severe’
N = 85

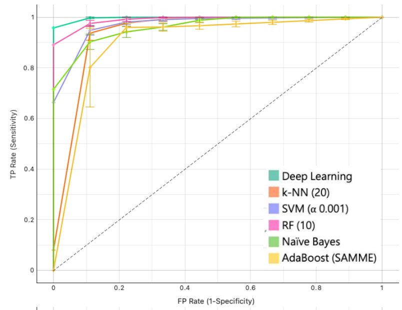
Model	Evaluation	
	AUC	F1
DL	0.982	0.595
k-NN	0.880	0.185
SVM	0.846	0.191
RF	0.953	0.346
Bayes	0.934	0.301
AdaBoost	0.703	0.439



Category: Anxiety
Class: ‘Extremely Severe’
N = 186

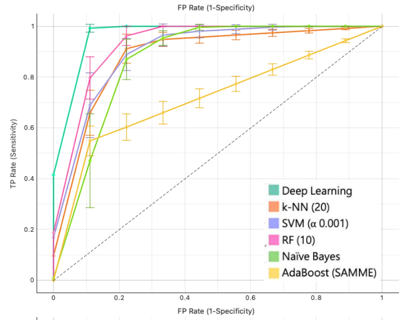
Model	Evaluation	
	AUC	F1
DL	0.998	0.903
k-NN	0.975	0.809
SVM	0.985	0.713
RF	0.992	0.830
Bayes	0.991	0.777
AdaBoost	0.960	0.809

Figure 2. The ROC analysis of anxiety emotional state



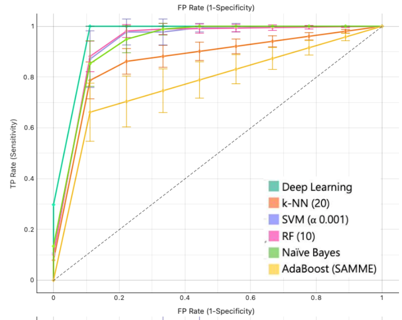
Category: Stress
Class: ‘Normal’
N = 1988

Model	Evaluation	
	AUC	F1
DL	0.998	0.897
k-NN	0.970	0.925
SVM	0.977	0.963
RF	0.992	0.961
Bayes	0.963	0.908
AdaBoost	0.909	0.958



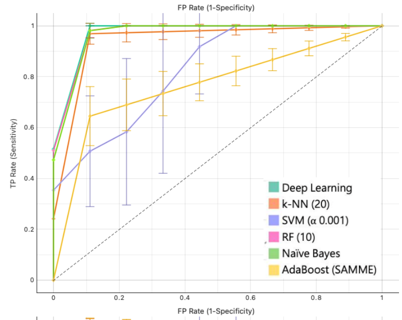
Category: Stress
Class: ‘Mild’
N = 261

Model	Evaluation	
	AUC	F1
DL	0.987	0.806
k-NN	0.886	0.202
SVM	0.906	0.566
RF	0.940	0.524
Bayes	0.865	0.459
AdaBoost	0.733	0.533



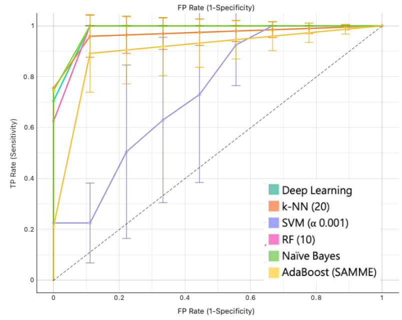
Category: Stress
Class: ‘Moderate’
N = 178

Model	Evaluation	
	AUC	F1
DL	0.987	0.773
k-NN	0.879	0.341
SVM	0.940	0.468
RF	0.940	0.478
Bayes	0.938	0.461
AdaBoost	0.791	0.480



Category: Stress
Class: ‘Severe’
N = 153

Model	Evaluation	
	AUC	F1
DL	0.994	0.827
k-NN	0.971	0.703
SVM	0.807	0.519
RF	0.978	0.694
Bayes	0.984	0.702
AdaBoost	0.795	0.646



Category: Stress
Class: ‘Extremely severe’
N = 44

Model	Evaluation	
	AUC	F1
DL	0.993	0.738
k-NN	0.975	0.762
SVM	0.728	0.370
RF	0.986	0.734
Bayes	0.997	0.763
AdaBoost	0.931	0.370

Figure 3. The ROC analysis of stress emotional state

From Fig.3, the ROC analysis in the ‘severe’ and ‘extremely severe’ classes also suffered from the lower sample size. The average AUC and F1 from all classes are 0.993, and 0.937, respectively.

The study shows that in a smaller sample class, false positives are increasing because the sample size within the subcategory is vary. To identify the most accurate model to classify the emotional states with the DASS-21 profile, using only AUC and CA might not be enough. In this study, we are taking the F1 score more intensely because of the varying sample size within each class. The mean F1 score visualization for each model is shown in Fig. 4.

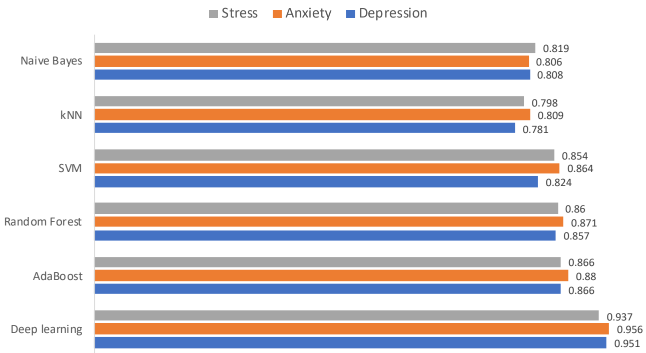


Figure 4. The mean F1 scores classified by emotional states.

TABLE III. Models Performance Evaluation

	$\overline{AUC}$	$\overline{CA}$	$\overline{F1}$	*Mean Precision*	*Mean Recall*
Depression					
Deep Learning	0.996	0.951	0.951	0.952	0.951
AdaBoost	0.891	0.866	0.866	0.866	0.866
Random Forest	0.972	0.865	0.857	0.855	0.865
SVM	0.942	0.838	0.824	0.867	0.838
kNN	0.945	0.808	0.781	0.776	0.808
Naive Bayes	0.951	0.792	0.808	0.837	0.792
Anxiety					
Deep Learning	0.996	0.955	0.956	0.957	0.955
AdaBoost	0.886	0.880	0.880	0.880	0.880
Random Forest	0.975	0.879	0.871	0.866	0.879
SVM	0.944	0.872	0.864	0.888	0.872
kNN	0.940	0.839	0.809	0.807	0.839
Naive Bayes	0.951	0.782	0.806	0.847	0.782
Stress					
Deep Learning	0.993	0.938	0.937	0.938	0.938
Random Forest	0.972	0.875	0.866	0.861	0.875
AdaBoost	0.838	0.860	0.860	0.861	0.860
SVM	0.900	0.854	0.854	0.886	0.854
kNN	0.938	0.835	0.798	0.793	0.835
Naive Bayes	0.942	0.795	0.819	0.864	0.795

From Table 3, the performance evaluation of six machine learning models—deep learning, AdaBoost, random forest, support vector machine (SVM), k-nearest neighbors (k-NN), and Naïve Bayes—were presented in classifying the emotional states of depression, anxiety, and stress. The performance metrics reported include the area under the curve (AUC), classification accuracy (CA), F1 score, mean precision, and mean recall.

Among all models, deep learning consistently outperformed the others across all three emotional state categories. It achieved the highest AUC (0.996 for depression, 0.996 for anxiety, and 0.993 for stress), indicating its superior ability to distinguish between severity levels. The classification accuracy for deep learning was also the highest, with 95.1% for depression, 95.5% for anxiety, and 93.8% for stress. The model's F1 score, a critical measure of precision and recall balance, similarly demonstrated high performance, with values of 0.951 for depression, 0.956 for anxiety, and 0.937 for stress.

Other models, such as AdaBoost and random forest, also performed relatively well but did not reach the same levels of accuracy or precision as deep learning. AdaBoost showed a lower AUC across all categories (0.891 for depression, 0.886 for anxiety, and 0.838 for stress), along with reduced classification accuracy and F1 scores. Random forest performed comparably better, with AUC values around 0.972 for depression, 0.975 for anxiety, and 0.972 for stress, but still fell short of deep learning in terms of overall precision and recall.

SVM, k-NN, and Naïve Bayes, while functional, demonstrated lower performance across most metrics, with k-NN and Naïve Bayes in particular showing reduced accuracy and F1 scores compared to the other models. This suggests that deep learning is the most reliable method for classifying emotional states when using the DASS-21 profile.

IV. Discussion

The findings of this study indicate that deep learning models outperformed other machine learning models in classifying the emotional states of depression, anxiety, and stress using the DASS-21 profile. The deep learning model consistently achieved high accuracy, precision, and AUC values across all emotional state categories, demonstrating its effectiveness in predicting the severity of mental health conditions, as well as highlighting the potential of using machine learning techniques to automate and improve mental health screening processes, particularly in non-clinical or early diagnostic settings where traditional manual assessments may be limited.

However, despite the strong performance of the deep learning model, the study also uncovered notable limitations that must be addressed. The model's performance declined in the "severe" and "extremely severe" categories, where sample sizes were smaller. This imbalance affected the model's F1 scores in these critical classes, potentially increasing the rate of false positives and limiting its reliability in identifying severe cases, which could compromise its clinical reliability in real-world applications where severe cases are often less common but critically important. Therefore, future work should explore methods to address class imbalance, such as using data augmentation, resampling techniques, or hybrid models that can enhance prediction accuracy in smaller sample groups.

Additionally, while this study focused on DASS-21, future iterations of the model could benefit from incorporating more diverse features, such as demographic data, psychosocial variables, or other clinical parameters, to further improve prediction performance. Integrating such features could make the model more robust and applicable to broader populations, ensuring it is more adaptable to various clinical contexts.

Overall, this study provides a promising framework for machine learning-based mental health assessment but also highlights areas for refinement to optimize the model's clinical utility.

V. Conclusion

This study demonstrates the potential of deep learning models to accurately and efficiently classify emotional states of depression, anxiety, and stress based on the DASS-21 profile. The results suggest that transitioning from traditional scoring-based assessments to machine learning-based models in clinical care could significantly enhance mental health screening and decision-making processes. While the deep learning model performed well overall, limitations were observed in its ability to accurately classify severe and extremely severe emotional states, particularly due to class imbalances in the dataset. Addressing these imbalances through techniques such as data augmentation and incorporating additional features could further improve the model's performance in critical cases.

Moreover, enhancing the model by integrating more diverse clinical and psychosocial parameters would make it more robust and adaptable to a wider range of populations and clinical settings. This study provides a foundation for the use of machine learning in mental health assessments, with the potential to improve both patient self-screening and clinical evaluations. Moving forward, further refinement of the model will be necessary to ensure its reliability in accurately identifying individuals across all severity levels, especially in real-world clinical applications.

Acknowledgment

Research supported by Unit of Excellence for Clinical Outcomes Research and Integration (UNICORN), University of Phayao, and Multimedia Data Analytics and Processing Research Unit, Chulalongkorn University, and the University of Phayao Counselling Center (UPCC) for providing the dataset.

References

[1] WHO, "COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide," *World Health Organization*, Mar. 02, 2022. https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide (accessed Feb. 06, 2023).

[2] Y. Li, A. Wang, Y. Wu, N. Han, and H. Huang, "Impact of the COVID-19 Pandemic on the Mental Health of College Students: A Systematic Review and Meta-Analysis," *Front Psychol*, vol. 12, Jul. 2021, doi: 10.3389/fpsyg.2021.669119.

[3] S. H. Lovibond and P. F. Lovibond, *Manual for the Depression Anxiety Stress Scales*. Psychology Foundation of Australia, 1996.

[4] J. D. Henry and J. R. Crawford, "The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample," *British Journal of Clinical Psychology*, vol. 44, no. 2, pp. 227–239, Jun. 2005, doi: 10.1348/014466505X29657.

[5] M. Makara-Studzińska, E. Tyburski, M. Załuski, K. Adamczyk, J. Mesterhazy, and A. Mesterhazy, "Confirmatory Factor Analysis of Three Versions of the Depression Anxiety Stress Scale (DASS-42, DASS-21, and DASS-12) in Polish Adults," *Front Psychiatry*, vol. 12, Jan. 2022, doi: 10.3389/fpsyt.2021.770532.

[6] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," *Procedia Comput Sci*, vol. 167, pp. 1258–1267, Jan. 2020, doi: 10.1016/J.PROCS.2020.03.442.

[7] P. Kumar, S. Garg, and A. Garg, "Assessment of Anxiety, Depression and Stress using Machine Learning Models," *Procedia Comput Sci*, vol. 171, pp. 1989–1998, 2020, doi: 10.1016/j.procs.2020.04.213.

[8] University of Phayao Counseling Center, "UP Mental Health Self-Assessment." http://165.22.255.52:8080/#/ (accessed Feb. 07, 2023).

[9] *Personal Data Protection Act, B.E. 2562 (2019)*, vol. 136 Chapter 69 Gor. Bangkok, 2019.

[10] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html